

## Research Article

---

# Effect of outgroup on phylogeny reconstruction: a case study of family Solanaceae

Ishrat Jamil\*, Syeda Qamarunnisa and Abid Azhar

The Karachi Institute of Biotechnology and Genetic Engineering (KIBGE), University of Karachi, Karachi-Pakistan

\*Corresponding author's email: [ishratjamil.kibge@gmail.com](mailto:ishratjamil.kibge@gmail.com)

### Citation

Ishrat Jamil, Syeda Qamarunnisa and Abid Azhar. Effect of outgroup on phylogeny reconstruction: a case study of family Solanaceae. Pure and Applied Biology. Vol. 8, Issue 4, pp2213-2227.

<http://dx.doi.org/10.19045/bspab.2019.80167>

---

Received: 06/02/2019

Revised: 10/07/2019

Accepted: 15/07/2019

Online First: 23/07/2019

---

### Abstract

In phylogenetic analysis, direction of ingroup evolution depends on outgroup, therefore, the determination of correct outgroup is very critical. In the current study, effect of different outgroups on tree topology within Solanaceae was examined using Bayesian analysis. Different outgroups were selected from closely related families of Solanaceae (Convolvulaceae, Hydrolaceae and Montiniaceae), distantly related families (Acanthaceae, Scrophulariaceae and Boraginaceae) and more distantly related group (Gymnospermic families; Cycadaceae and Pinaceae). Single taxa from each family, multiple taxa from a single genus and multiple taxa from different genera of the same sister family were selected to evaluate their resolving power. The single taxa of a genus belong to Convolvulaceae produced more consistent result as compared to the multiple taxa of the single genus. In addition, single taxa produced similar tree topology as produced by multiple taxa of different genera. Among single taxa, *Evolvulus pilosus*, observed to be the best outgroup because it resolved the associations between genus, tribes, and subfamilies within Solanaceae followed by *Humbertiama dagascariensis*. The lineages established by the successful outgroups suggest the evolutionary pattern within Solanaceae is Cestroioideae to Petunoideae to Solanoideae. Distantly related outgroup was unable to resolve the tree topology. It may be suggested that for lower level taxonomical relationships of Solanaceae, *Evolvulus pilosus* could be used as outgroup to infer true associations.

**Keywords:** Ingroup; Outgroup; Phylogenetics; *rbcL*; Solanaceae

### Introduction

Molecular phylogenetic infers the hypothesis of evolutionary relationships in the organisms or group of organisms by means of phylogenetic tree. The topology of the tree, which is based on hypothesis of relationships, is effected by several assumptions and factors including sample size, selection of coding or noncoding region of the genome, sequence length of the selected region, alignment methodology, optimality criterion, treatment of data, parameter values and assumptions of each phylogenetic inference methods [1-4]. Selection of outgroup is among those explicit and implicit factors that affects the

topology of the phylogenetic tree. However, there is no extensive research presented on its use in the reconstruction of the phylogenetic tree. It usually been selected randomly or on the basis of vague relationships between in group and outgroup taxa [1, 3, 5-8]. An out group is used to root the unrooted cluster. It infers the theoretical ancestral origin and provides the array of time of all subsequently evolutionary events led to diverging the sequence, because the root of a tree is first or deepest split [1, 5-9]. An out group is necessary for phylogenetic analysis to understand the evolution of the group under study (ingroup). As the direction of ingroup evolution depends on

the out group, therefore, the selection of correct outgroup is very critical. An incorrect outgroup may mislead the inference of hypothetical taxonomic relationships and character evolution. Therefore, in any phylogenetic analysis the determination of correct root location with accuracy is very important. Incorrectly rooted tree may provide erroneous phylogenetic signals that may lead to the “random outgroup effect” and Long Branch Attraction (LBA). To avoid such errors, outgroup should be from outside the ingroup but closely related to the group of study. Usually sister group of the ingroup is preferably chosen as outgroup. Moreover, outgroup contains multiple taxa has been given more consideration because of better inference [2, 3, 6-13]. However, it is not always necessary that the closely related sister taxa provide correct hypothesis of phylogenetic signals especially in case of the relatively high evolutionary rate or if the only single taxon, as outgroup, is present in the sister group [1, 5, 8].

In the present work, effort was made to study the possible effects of different outgroups to infer the subfamilial relationships in Solanaceae.

## Materials and methods

### Ingroup plant material

Plant samples for ingroup were collected from different areas of Karachi during their flowering season and identified with the help of key characters present in flora of Pakistan [14] and other floras [15, 16]. Herbarium sheets of the identified species were prepared and submitted to Karachi University Herbarium. Their names, locations, voucher number and GenBank accession numbers are given in (Table 1).

### Outgroup selection

Most of the outgroup species sequences were retrieved from GenBank. Outgroup was selected as closely related, distantly related and more distantly related members with respect to Solanaceae. An outgroup *Schizanthus pinnatus* was also selected randomly from the ingroup (Solanaceae, subfamily Schizanthoideae). Different members from sister families of Solanaceae (Convolvulaceae, Hydroclaceae, Montineaceae) were chosen as outgroup. Different species of same genera were also checked for their potency to be selected as outgroup. Taxa from distantly related families like Acanthaceae, Scrophularaceae and

Boraginaceae were also tested. More distantly related taxa belong to Gymnosperm; *Cycas revoluta* (Cycadaceae) and *Pinus gerardiana* (Pinaceae) were also checked for their resolution power in Solanaceae phylogenetic tree. Multiple taxa that belong to same and different genera were also taken into consideration to reconstruct the phylogeny of Solanaceae. A list of selected outgroup species with their classification and accession numbers is given in (Table 2) while their sequence length and GC content is mentioned in (Table 3).

### DNA extraction and PCR

A modified Cetyl trimethyl ammonium bromide (CTAB) method [17] was used to extract the total genomic DNA from the fresh leaf samples. By using primer3 software version 0.4.0 [18] family specific primers were designed for the coding region (*rbcL*) of chloroplast genome. Polymerase chain reaction (PCR) was performed in thermal cycler to amplify the *rbcL* gene. Amplification of the *rbcL* gene was carried out in 20 µL reaction volume containing 1X PCR buffer, 2 mM MgCl<sub>2</sub>, 0.2 mM dNTPs, 0.3 µM of each primer (Forward, 5'-TTCCGATTCACCGGATCTTA-3' and reverse, 5'-ACTCATTGTTATAGCTGGAT-3'), 1.25 units of DNA *Taq* Polymerase (MOLEQULE-ON, New Zealand), 100 ng DNA template and an appropriate amount of Milli-Q water. The PCR conditions were set as: initial denaturation of double stranded DNA at 94°C for 4 minute, 35 cycles of 94°C for 30 second, primer annealing at 54°C for 1 minute, extension at 72°C for 1.5 minutes and after 35 cycles final extension at 72°C for 7 minutes. Amplified products were observed on 1% agarose gel and purified using column based PCR purification kit (MOLEQULE-ON, NZ). Purified products were sent for direct sequencing at MOLEQULE-ON (NZ).

### Sequence Editing and Alignment

Basic Local Alignment Search Tool, Blastn [19] was performed to analyze each nucleotide sequence. Open reading frame, ORF [20] was performed to observe the putative conserved domain of the *rbcL* gene. The discrepancy present in nucleotide sequence was resolved by aligning forward and reverse primer sequences using Multalin software [21]. Nucleotide sequences were translated into protein using online available ExPASy translate tool [22]. The refined sequences were submitted to NCBI

(National Center for Biotechnology Information) Genbank to get their specific accession numbers.

### Phylogenetic tree reconstruction

Phylogenetic tree was reconstructed by using each chosen outgroup. The model based Bayesian inference method was used to infer the relationship within Solanaceae. The best substitution model for almost all the combinations was found to be Hasegawa, Kishino and Yano (HKY) as determined by JModelTest 2.1.3 software [23]. The program BEAUTi [24] was executed to generate the xml file which was further executed in BEAST. MCMC chain length was set as 800,000 generations. Tree was screened at every 10,000 generations. The burn-in value was set as 40.

### Tree topology

All phylogenetic trees were carefully analyzed in order to examine the ability of different outgroups in resolving infrafamilial relationships within Solanaceae. Tree topology for every tree was monitored. Lineages and posterior probability value were taken into consideration for elucidating best outgroup species.

### Results

The effect of different outgroups on the reconstruction of Solanaceae phylogeny was observed with the help of Bayesian approach. Different members (29) of family Solanaceae were taken to be included as ingroup, out of which 16 were collected and processed for downstream procedures while sequences of remaining 5 members were taken from online sources (GenBank). 40 species from different genera of closely and distantly related families were chosen as outgroup, out of which only 10 species from closely related genera, result in resolving established subfamilial relationships within Solanaceae including one Solanaceae member *Schizanthus*. Sequence length and GC content of most of the outgroup species was constant as shown in (Table 3). Description of subfamilies, their lineage and different tribes as inferred by different successful outgroup species is given in (Table 4). Posterior probability (PP) value was used to signify the ability of each successful species to be chosen as outgroup in Solanaceae studies where below family level relationships are to be determined. Furthermore, tribal connections were also evaluated as mentioned in (Table 5). It was observed that five

species of *Cuscuta*, one species of *Convolvulus*, two species of *Evolvulus* and one species of *Humbertia* were able to resolve the true topology of Solanaceae tree. One ingroup species *Schizanthus pinnatus* also resolved associations within Solanaceae. All potential outgroup species showed different branches of three subfamilies (Solanoideae, Petunoideae, Cestroideae), which further sub-divided into seven tribes (*Solaneae*, *Capsiceae*, *Physleae*, *Datureae*, *Lycieae*, *Hyoscymeeae*, *Cestreae*). As *Petunia* is represented by only one member therefore a separate branch protruding tribe *Petuneae* can also be seen. Evolutionary lineage of subfamilies was observed to be as Cestroideae to Petunoideae to Solanoideae by most of the outgroup species as shown in (Table 4) except three *Cuscuta* species (*Cuscuta gronavi*, *Cuscuta indecora* and *Cuscuta platyloba*), which displayed the ancestry of subfamilies as Petunoideae to Cestroideae to Solanoideae. *Evolvulus pilosus* was seen to be best outgroup species among all other analyzed based on the tree topology and supportive probability value (Figure 1). The tree supports three subfamilies, eight tribes and their associations. All genera showed monophyletic clades and species of a represented genus are supported by strong probability value. Mostly the strong probability value (within the range of 0.9 - 1) is seen to support the different tribes and subfamilies. In (Figure 2) displays different phylogenetic trees with successful outgroup that showed a resolved topology of tree. Only four trees are shown here to represent the topology of resolved trees by different outgroups. Most of the predictions regarding phylogenetic association in Solanaceae below family level are similar that are shown in (Table 4 & 5). Some of the unresolved topologies are also shown in (Figure 3) which demonstrate why some outgroup are not to be chosen in phylogenetic reconstruction. Almost all outgroups used in this study maintain the monophyletic status of genera however their influence to establish subfamilial and tribal relationships were different. Multiple taxa from same and different genera were also used to reconstruct the phylogeny of Solanaceae. It was observed that when taxa from different genera were used (Figure 4), a resolved topology was perceived. The tree showed the similar topology as observed by single outgroup member.

However when the different members of same genus were used, an unresolved phylogenetic tree

was observed that only established the monophyly of genus (Figure 5).

**Table 1. List of ingroup species with their location, voucher numbers and GenBank accession numbers**

S. #	Genus	Species Name	Locality	Voucher #	Accession #
1	Solanum	<i>Solanum nigrum</i>	Campus, University of Karachi	G. H. No. 86479	JX996056
2		<i>Solanum tuberosum</i>	Gulshan e Maymar, Karachi	G.H. No. 86582	KJ652187
3		<i>Solanum surattense</i>	Campus, University of Karachi	G. H. No. 86480	JX996057
4		<i>Solanum melongena</i>	Campus, University of Karachi	G. H. No. 86485	JX675575
5		<i>Solanum forskalii</i> Dunal	Safari park, Karachi	G. H. No. 86534	KJ652200
6		<i>Lycopersicon esculentum</i> L.	Center for Plant Conservation, University of Karachi	G. H. No. 86481	KJ652188
7		<i>Solanum torvum</i>			KC535809.1
8		<i>Solanum dulcamara</i>			HQ590279.1
9		<i>Solanum anguivi</i>			JX511989.1
10		<i>Solanum seafortheanum</i>			
11	Capsicum	<i>Capsicum annuum</i> L.	NARC, Islamabad	G.H. No. 86538	KJ652190
12		<i>Capsicum frutescens</i> L.	Shah Faisal colony, Karachi	G. H. No. 86480	JX996068
13	Physalis	<i>Physalis divaricata</i> D.Don	Campus, University of Karachi	G. H. No. 86474	JX996060
14		<i>Physalis alkekengi</i>			U08617.1
15		<i>Physalis peruviana</i>			FJ914181.1
16	Withania	<i>Withania coagulans</i> (Stocks) Dun	Center for Plant Conservation, University of Karachi	G. H. No. 86484	KC291705
17		<i>Withania somnifera</i> (L.) Dunal	Campus, University of Karachi	G. H. No. 86476	JX996055
18	Lycium	<i>Lycium edgeworthii</i> (Dunal)	Center for Plant Conservation, University of Karachi	G. H. No. 86533	KJ652191
19		<i>Lycium chinense</i>			FJ914171.1
20		<i>Lycium ruthenicum</i>			JF942340.1
		<i>Lycium shawii</i>			KU757338.1
21	Datura	<i>Datura stramonium</i> L.	Center for Plant Conservation, University of Karachi	G. H. No. 86475	JX996058
22		<i>Datura inoxia</i> Miller	Campus, University of Karachi	G. H. No. 86478	JX996059
24	Cestrum	<i>Cestrum nocturnum</i> L.	Center for Plant Conservation, University of Karachi	G. H. No. 86535	KJ652189
25		<i>Cestrum diurnum</i> L.	Center for Plant Conservation, University of Karachi	G. H. No. 86532	KJ653359
26		<i>Cestrum aurantiacum</i>			JX856311.1
27		<i>Atropa belladonna</i>			U08609.1
28		<i>Hyoscyamus niger</i>			HQ216125.1
29		<i>Petunia axillaris</i>			HQ384915.1

**Table 2. List of selected outgroup species with their classification and accession numbers**

Group	Order	Family	Subfamily	Genus	Species	Accession #
Angiosperm	Solanales	Solanaceae	Schizanthoideae	Schizanthus	<i>Schizanthus pinnatus</i>	U08619.1
		Convolvulaceae	Convolvuloidea	Convolvulus	<i>Convolvulus arvensis</i>	AY100993.1
					<i>Convolvulus kotschyanus</i>	KT596055
					<i>Convolvulus glomeratus</i>	KT596047
					<i>Convolvulus scindicus</i>	KT596046
					<i>Convolvulus prostrates</i>	KT596045
				Cuscuta	<i>Cuscuta rostrata</i>	EU330263.1
					<i>Cuscuta platyloba</i>	AJ320209.1
					<i>Cuscuta indecora</i>	EU330274.1
					<i>Cuscuta europaea</i>	AY101060.1
					<i>Cuscuta gronavi</i>	AJ320207.2
				Ipomoea	<i>Ipomoea batata</i>	AY100962.1
					<i>Ipomoea cairica</i>	KJ652199
					<i>Ipomoea indica</i>	KJ773595.1
					<i>Ipomoea purpurea</i>	NC_009808.1
					<i>Ipomoea pandurata</i>	KJ773596.1
				<i>Dinetus</i>	<i>Dinetus truncates</i>	AY101053.1
				<i>Tridynamia</i>	<i>Tridynamia megalantha</i>	AY101054.1
				<i>Humbertia</i>	<i>Humbertia madagascariensis</i>	AY101062.1
				Evolvulus	<i>Evolvulus nuttallianus</i>	KT178137.1
					<i>Evolvulus sericeus</i>	KJ773505.1
					<i>Evolvulus pilosus</i>	HQ384917.1
					<i>Evolvulus glomeratus</i>	AY101012.1
				Jacquemontia	<i>Jacquemontia reclinata</i>	AY101040.1
					<i>Jacquemontia blanchetii</i>	AY101039.1
					<i>Jacquemontia sandwicensis</i>	AY101038.1
					<i>Jacquemontia tamnifolia</i>	AY101037.1
<i>Jacquemontia pentantha</i>	AY101036.1					
Calystegia	<i>Calystegia sepium</i>	AY100992.1				
	<i>Calystegia macrostegia</i>	AY100991.1				
	<i>Calystegia pulchra</i>	KF997266.1				
	<i>Calystegia silvatica</i>	KF997298.1				
	<i>Calystegia soldanella</i>	KT626673.1				
Gymnosperm	Solanales	Hydroleaceae	Hydrophyllodeae	<i>Hydrolea</i>	<i>Hydrolea ovate</i>	L14293.1
	Solanales	Montiniaceae	-----	<i>Montinia</i>	<i>Montinia caryophyllacea</i>	L11194.2
	Lamiales	Acanthaceae	Acanthoideae	<i>Thunbergia</i>	<i>Thunbergia alata</i>	HQ384878.1
		Scrophulariaceae	Scrophularioidea	<i>Scrophularia</i>	<i>Scrophularia californica</i>	HQ384892.1
	Brassicales	Brassicaceae	.....	<i>Arabidopsis</i>	<i>Arabidopsis thaliana</i>	AB917053.1
Gymnosperm	Cycadales	Sycadaceae	-----	<i>Cycus</i>	<i>Cycus revoluta</i>	AY056556.1
	Pinales	Pinaceae	Pinoideae	<i>Pinus</i>	<i>Pinus gerardiana</i>	AY115762.1

**Table 3. Sequence length and GC content of included outgroup species**

S. No.	Name of species	GC content %	DNA length
1	<i>Schizanthus pinnatus</i>	43.6	1408
2	<i>Humbertiana dagascariensis</i>	44.1	1434
3	<i>Cuscuta platyloba</i>	41.8	1399
4	<i>Cuscuta indecora</i>	41.9	1307
5	<i>Cuscuta rostrata</i>	42.6	1324
6	<i>Cuscuta europaea</i>	43.8	1314
7	<i>Cuscuta gronavi</i>	39.3	1931
8	<i>Convolvulus arvensis</i>	44.7	1428
9	<i>Convolvulus kotschyanus</i>	44.7	606
10	<i>Convolvulus glomeratus</i>	43.4	970
11	<i>Convolvulus scindicus</i>	43.4	1020
12	<i>Convolvulus prostratus</i>	42.9	976
13	<i>Ipomoea batata</i>	44.2	1333
14	<i>Ipomoea cairica</i>	44.3	1443
15	<i>Ipomoea indica</i>	44.6	1323
16	<i>Ipomoea pandurata</i>	45.1	1309
17	<i>Ipomoea purpureae</i>	43.8	1443
18	<i>Evolvulus nuttallianus</i>	44.7	1461
19	<i>Evolvulus sericeus</i>	44.0	1256
20	<i>Evolvulus pilosus</i>	44.1	1308
21	<i>Evolvulus glomeratus</i>	44.2	1408
22	<i>Jacquemontia reclinata</i>	44.0	1351
23	<i>Jacquemontia blanchetii</i>	44.6	1301
24	<i>Jacquemontia sandwicensis</i>	44.1	1340
25	<i>Jacquemontia tamnifolia</i>	44.5	1402
26	<i>Jacquemontia pentantha</i>	44.4	1401
27	<i>Calystegia sepium</i>	44.7	1424
28	<i>Calystegia macrostegia</i>	44.6	1398
29	<i>Calystegia pulchra</i>	41.5	1371
30	<i>Calystegia silvatica</i>	42.7	1383
31	<i>Calystegia soldanella</i>	44.8	1304
32	<i>Arabidopsis thaliana</i>	44.1	1326
36	<i>Montinia caryophyllacea</i>	43.4	1422
33	<i>Dinetus truncatus</i>	44.9	1392
35	<i>Tridynamia megalantha</i>	45.5	1370
34	<i>Hydrolea ovata</i>	42.1	1414
37	<i>Thunbergia alata</i>	44.4	1461
38	<i>Scrophularia californica</i>	42.9	1460
39	<i>Cycus revoluta</i>	44.5	1402
40	<i>Pinus gerardiana</i>	45.0	1413

**Table 4. Successful outgroup species and their resolution at tribal level within Solanaceae**

S. No.	Name of outgroup species	Subfamily Lineage	Tribes support with PP value (range 0-1)						
			Solaneae	Capsiceae	Physaleae	Datureae	Lycieae	Hyoscyameae	Cestreae
1	<i>Convolvulus arvensis</i>	CE→PE→SL	0.87	1	0.91	1	0.99	1	0.98
2	<i>Cuscuta europaea</i>	CE→PE→SL	0.97	1	1	1	1	1	0.98
3	<i>Cuscuta gronavi</i>	PE→CE→SL	0.99	1	0.98	1	1	1	1
4	<i>Cuscuta indecora</i>	PE→CE→SL	0.98	1	0.99	1	1	1	0.99
5	<i>Cuscuta platyloba</i>	PE→CE→SL	0.99	1	1	1	1	1	0.99
6	<i>Cuscuta rostrata</i>	CE→PE→SL	0.99	1	1	1	1	1	0.94
7	<i>Evolvulus glomeratus</i>	CE→PE→SL	0.95	1	0.95	1	0.98	1	0.89
8	<i>Evolvulus pilosus</i>	CE→PE→SL	0.98	1	1	1	1	1	0.91
9	<i>Humbertiama dagascariensis</i>	CE→PE→SL	0.99	1	1	1	1	1	0.91
10	<i>Schizanthus pinnatus</i>	CE→PE→SL	0.96	1	0.97	1	1	1	0.98

Abbreviations defined: SL: Solanoideae, PE: Petunoideae, CE: Cestroideae

**Table 5. Tribal associations as inferred by different outgroup species**

S. No.	Name of outgroup species	Tribal associations supported by PP value		
		Capsiceae - Datureae	(Capsiceae – Datureae)- Solaneae	Hyoscyameae- Lycieae
1	<i>Convolvulus arvensis</i>	0.42	0.57	0.85
2	<i>Cuscuta europaea</i>	0.26	0.37	0.52
3	<i>Cuscuta gronavi</i>	0.34	0.51	0.66
4	<i>Cuscuta indecora</i>	0.28	0.38	0.41
5	<i>Cuscuta platyloba</i>	0.34	0.56	0.65
6	<i>Cuscuta rostrata</i>	0.4	0.59	0.84
7	<i>Evolvulus glomeratus</i>	0.42	0.65	0.79
8	<i>Evolvulus pilosus</i>	0.4	0.65	0.69
9	<i>Humbertiama dagascariensis</i>	0.33	0.49	0.8
10	<i>Schizanthus pinnatus</i>	0.48	0.72	0.96

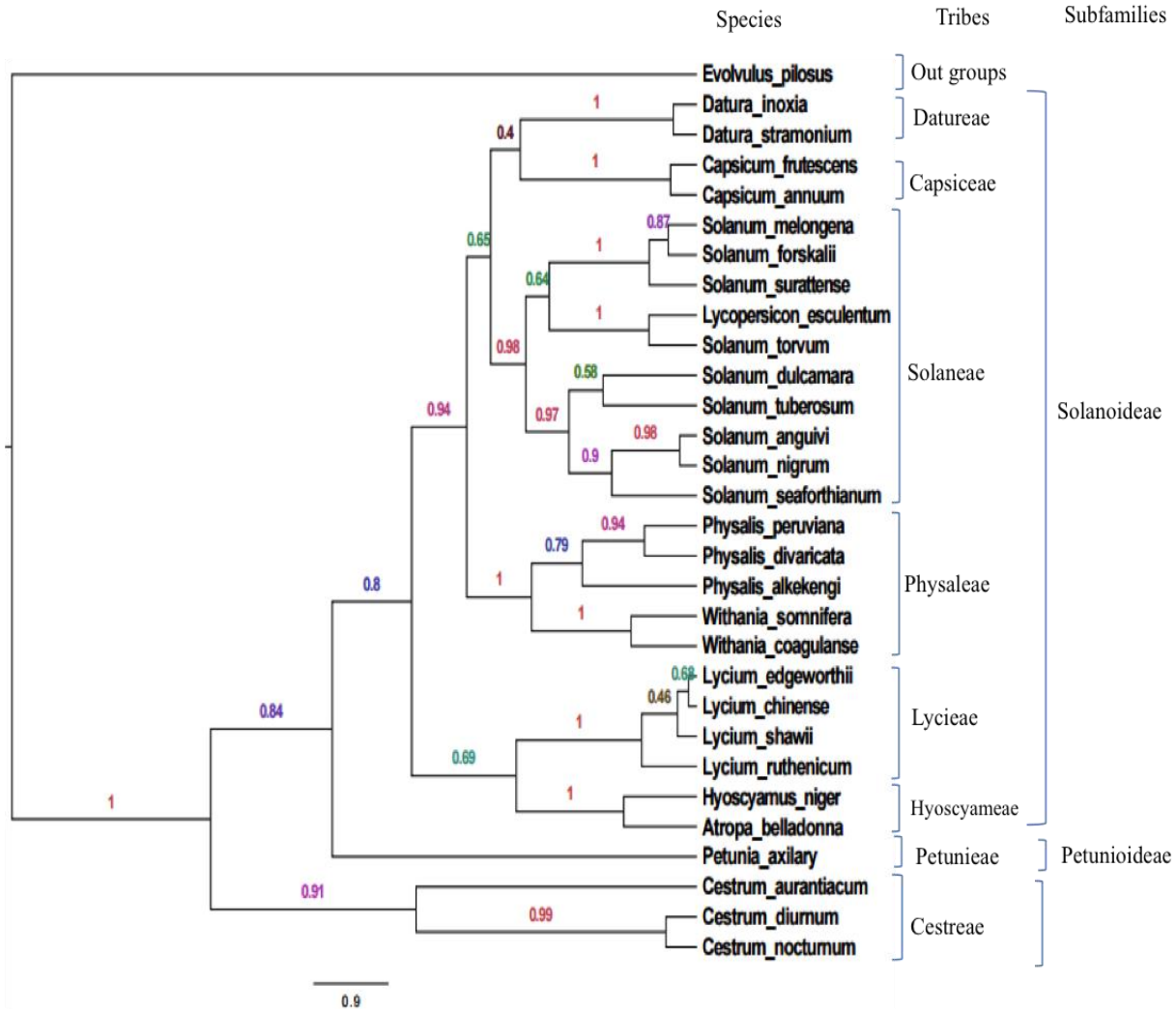
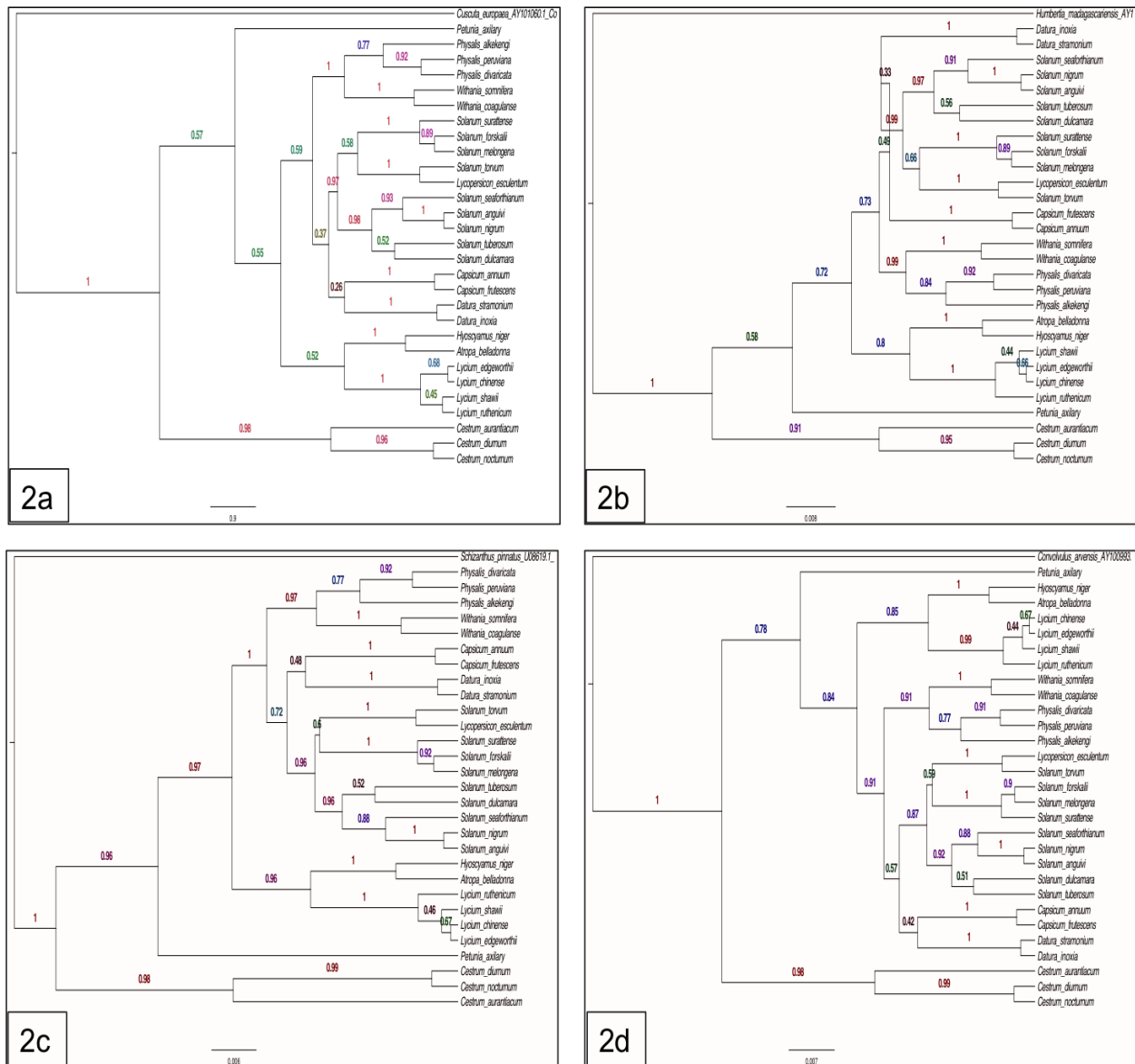
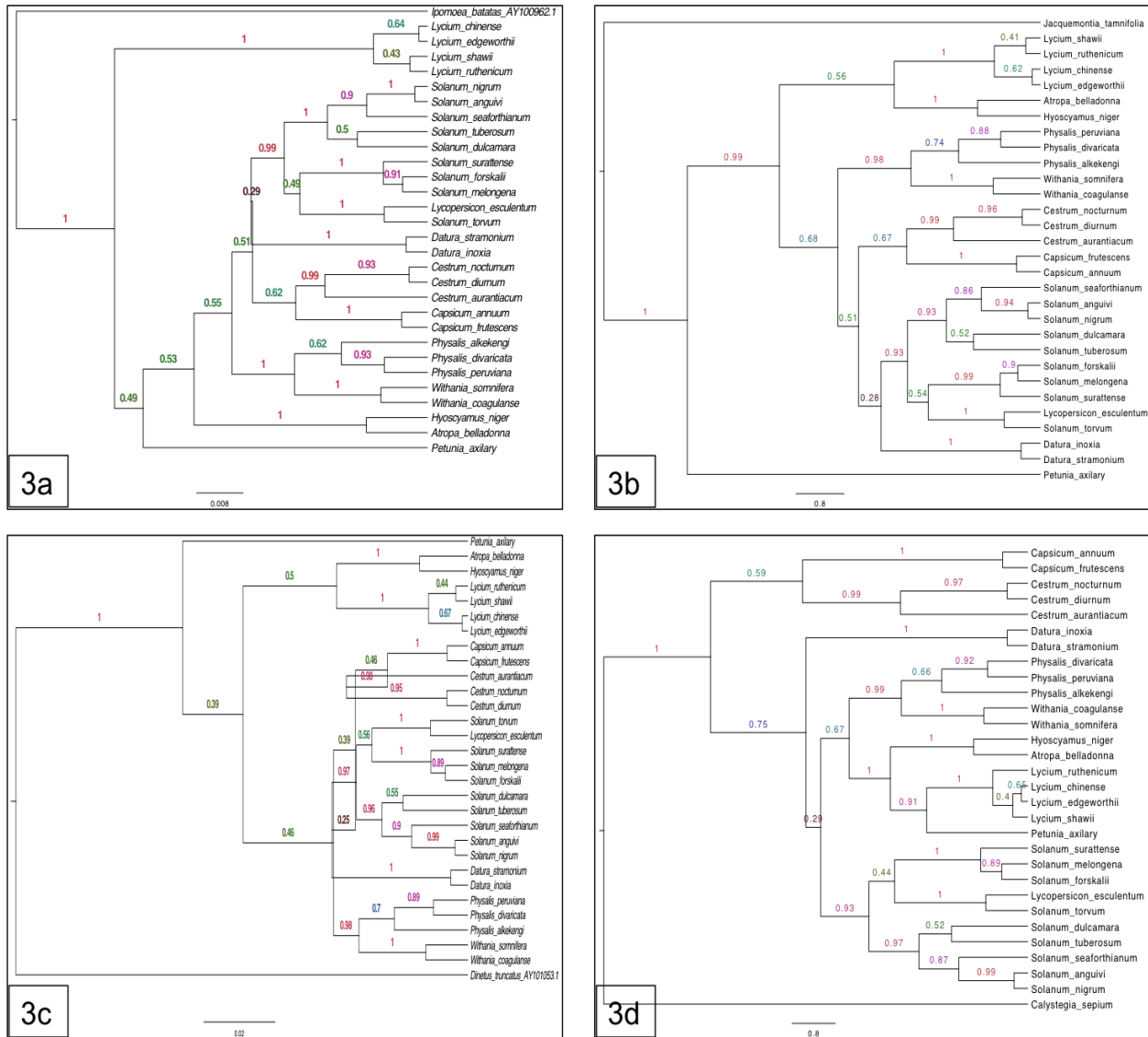


Figure 1. A resolved phylogenetic tree of Solanaceae as inferred by Bayesian inference using *rbcL* chloroplastic marker. *Evolvulus pilosus* was used as outgroup. Numbers above branches signifying posterior probabilities

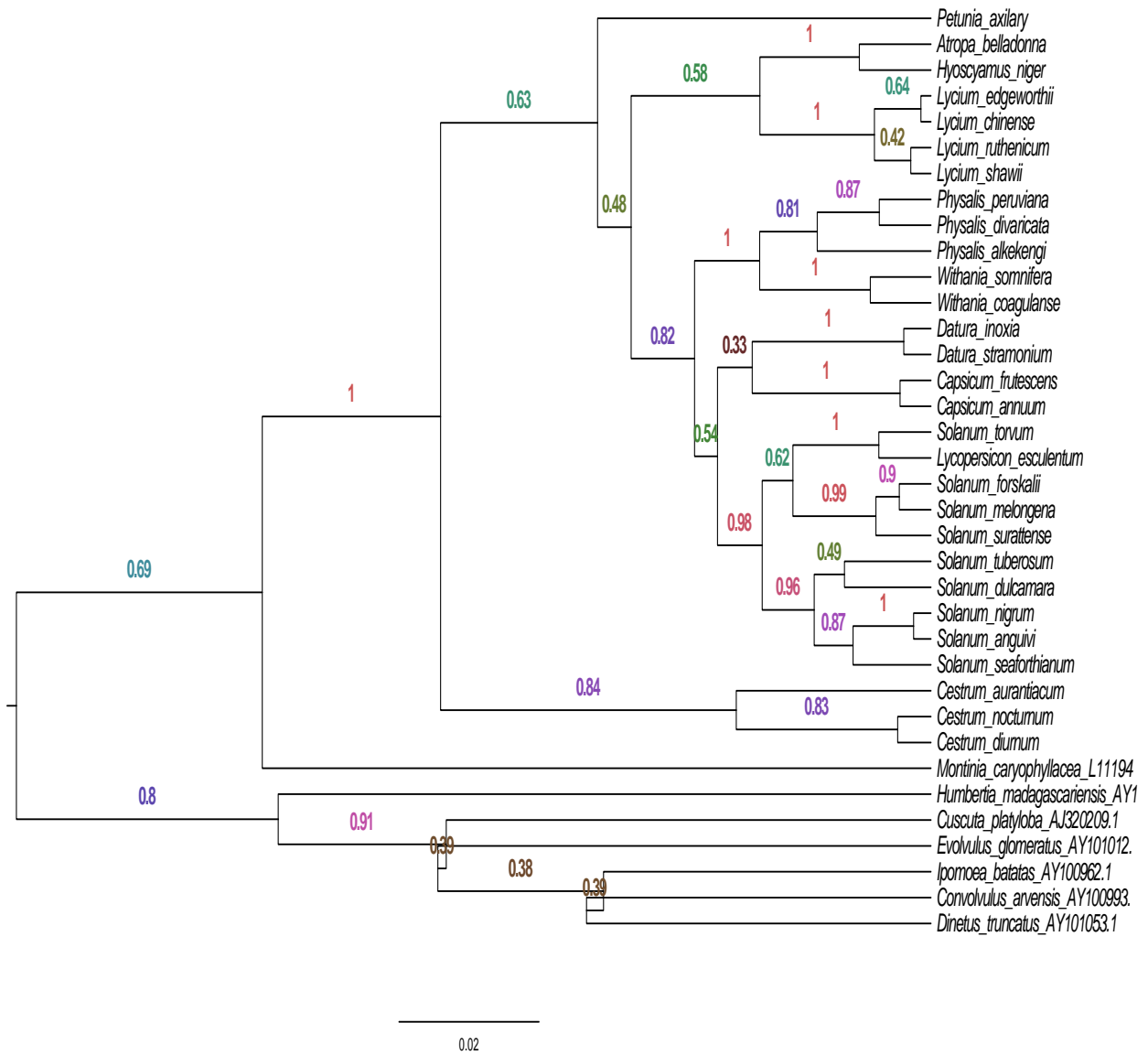




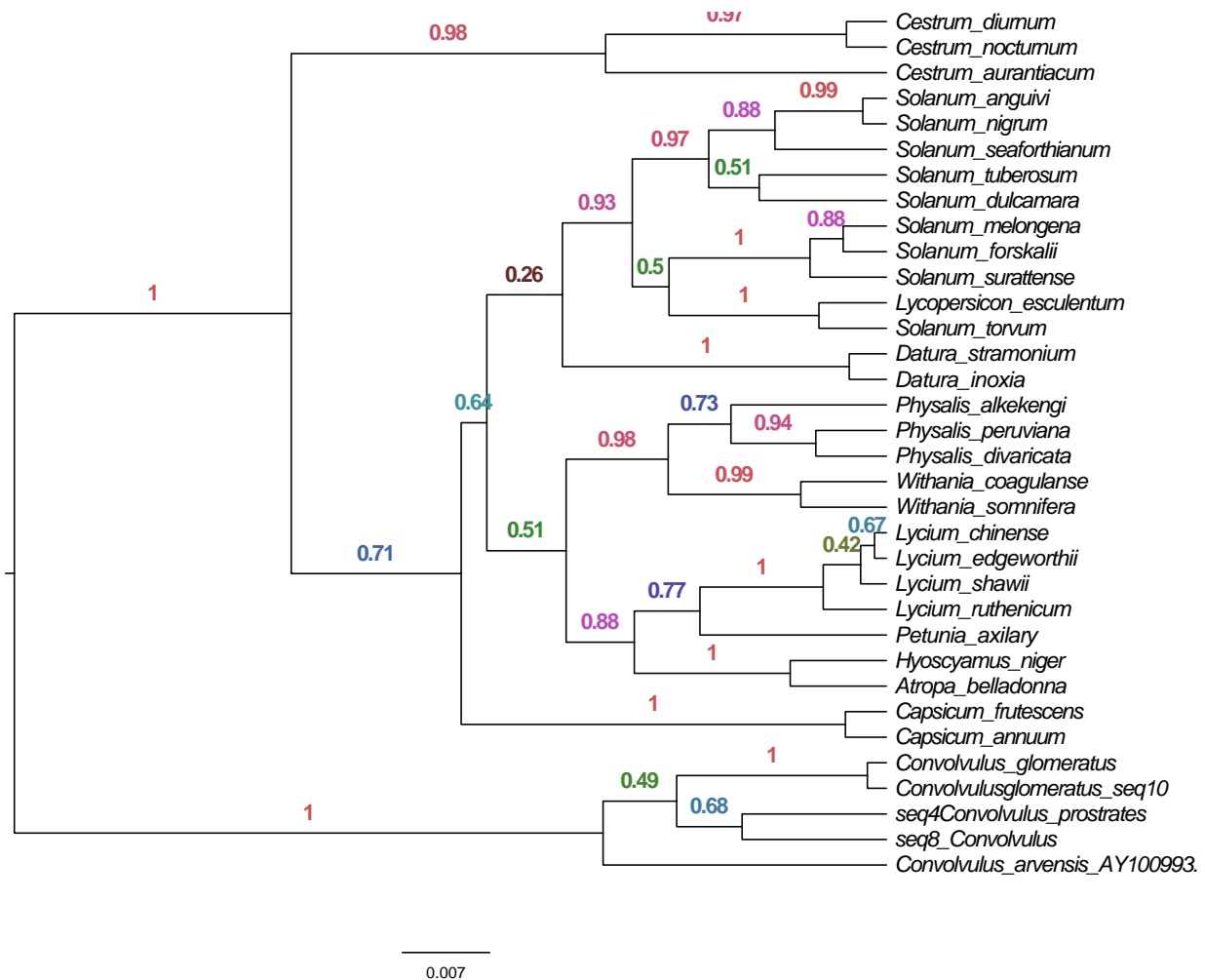
**Figure 2. Resolved phylogenetic trees of Solanaceae as inferred by Bayesian inference using *rbcL* chloroplastic marker. (2a) *Cuscuta europea* (2b) *Humbertiana dagascariensis* (2c) *Schizanthus pinnatus* (2d) *Convolvulus arvensis* were used as outgroup. Numbers above branches showing posterior probabilities**



**Figure 3. Unresolved phylogenetic trees of Solanaceae as inferred by Bayesian inference using *rbcL* chloroplastic marker. (3a) *Ipomoea batata* (3b) *Jacquemontia tamnifolia* (3c) *Dinetus truncatus* (3d) *Calystegia sepium* were used as outgroup. Numbers above branches showing posterior probabilities**



**Figure 4. A resolved phylogenetic tree of Solanaceae as inferred by Bayesian inference using *rbcL* chloroplast marker. Multiple taxa from different genera were used as outgroup. Numbers above branches showing posterior probabilities**



**Figure 5. An unresoluted phylogenetic tree of Solanaceae as inferred by Bayesian inference using *rbcL* chloroplastic marker. Multiple taxa from same genus *Convolvulus* were used as outgroup. Numbers above branches showing posterior probabilities**

**Discussion**

The choice of outgroup species in phylogenetics is very critical. An incorrect outgroup may mislead the inference of hypothetical taxonomic relationships and character evolution. In the current study, effort was made to highlight the effect of different outgroup in Solanaceae phylogeny. Outgroup taxa from the closely related family Convolvulaceae and distantly related families Montiniaceae, Acanthaceae and Scrophulariaceae were included in the study. Taxa from Gymnospermic families like Cycadaceae and Pinaceae were also taken into consideration to observe their effect.

Phylogenetic relationships of the Solanaceae were examined using chloroplastic gene, *rbcL* for 29 species. In Pakistan, Solanaceae is represented by 52 species [14], out of which 29 have been presented in the current study. Most of the ingroup species were freshly collected and used in further downstream procedures however, for outgroup taxa, mostly the sequences were retrieved from online available databases. The chloroplastic *rbcL* gene was selected to infer the subfamilial relationships within Solanaceae as this gene has proven ability to resolve the associations at various taxonomical levels [25-29]. The selection of outgroup in phylogenetic

reconstruction is very important and critical, as the wrong choice can mislead the associations at each taxonomic level. Previously, different studies have demonstrated the usefulness of single or multiple taxa as outgroup and their effect on tree topology and branch length [11, 30-33]. Outgroups were selected from different genus of sister family Convolvulaceae. From a particular genus, 5 species were chosen as outgroup. Represented members of other close families like Hydroleaceae and Montiniaceae were also chosen. Members from other orders like Lamiales, Brassicales and other Gymnospermic members were also evaluated. In the present work, almost all the reconstructed trees established a monophyletic branch for one genus species however, many of the subfamilies and tribes within the family do not form monophyletic clades by using different outgroup taxa. Sequence length and GC content of plant species used in phylogenetics play very significant role [34-36], care was taken while choosing outgroup species for the current work for these parameters. Out of 40 species, 10 outgroup species resolved the tree topology that was congruent with the already established relationships at subfamilial level [37-39]. Among those, *Evolvulus pilosus* may be said the best outgroup because it resolved the associations between genus, tribes, and subfamilies within Solanaceae that was already defined by previous morphological and molecular studies. The lineages established by the most successful outgroups suggest the evolutionary pattern within Solanaceae is Cestroideae to Petunoideae to Solanoideae. This hypothesis is also corresponding to the previous studies [38]. Three members of *Cuscuta* genus, *C. gronavi*, *C. indecora* and *C. platyloba* suggest that Cestroideae is evolved from Petunoideae, however, no evidence has been reported yet to support this hypothesis. The Bayesian inference method uses a posterior probability (PP) value to support the existence of a relationship between organisms [40-42], therefore the high probability value was considered to truly define the connections within Solanaceae. The 10 outgroup species presented in Table 4 resolved the tree topology in such a manner that the PP value was at the strongest side (~0.9 - 1). Other findings also suggested the significance of probability value in

character based phylogenetic methods [41]. The ten successful species in the current work represented 7 tribes in the Solanaceae for the included species. A single species tribe Petuneae can also be observed in those analyses. The existence of these tribes are in agreement with the previously reports [37-39]. Furthermore, tribal associations were also assessed in the light of earlier findings. This work also observed the same arrangement of tribal connections in the 10 outgroup species tree where Capsiceae closely related to Datoreae and to this clade Solaneae is associated while Hyoscymeeae connected to Lycieae.

Various combinations of outgroups were also assessed for their suitability and effects on tree topology were examined. It was observed that the tree reconstructed with various combination of outgroup species from different closely related genera was able to resolve the relationships. The tree that uses multiple outgroup from same genus did not establish the proven associations within Solanaceae. Previous reports observed the same findings on different groups. However, it can be said that the capability of single outgroup is found to be as strong as the combination of outgroup species.

### Conclusion

The present study established that every taxon has different ability to resolve the phylogenetic relationship for a defined ingroup species even if it is taken from the same genus. This competency of outgroup species depends mainly on its genetic code. The closeness of the outgroup member with the ingroup taxa both morphologically and genetically may be use to define the taxonomical connections. The capability can also be determined by strong probability or bootstrap value. However, ingroup topology was found to be sensitive to outgroup choice and increasing taxon sampling within the Solanaceae can result in more robust phylogenies.

### Authors' contributions

Conceived and designed the experiments: I Jamil & S Qamarunnisa, Performed the experiments: I Jamil, Analyzed the data: I Jamil & S Qamarunnisa, Contributed materials/ analysis/ tools: A Azhar, Wrote the paper: I Jamil & S Qamarunnisa.

## References

1. Smith AB (1994). Rooting molecular trees: problems and strategies. *Biol J Linnean Soc* 51: 279-292.
2. Dalevi D, Hugenholtz P & Blackall LL (2001). A multiple-outgroup approach to resolving division-level phylogenetic relationships using 16S rDNA data. *Int J Syst Evol Microbiol* 51: 385-391.
3. Cameron SL, Miller KB, D' Haese CA, Whiting MF & Barker SC (2004). Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entonatha, Insect and Crustacea Sensu lato (Arthropoda). *Cladistics* 20: 534-557.
4. Ware JL, Litman J, Klass KD & Spearman LA (2008). Relationships among the major lineages Dictyoptera: the effect of outgroup selection on dictyopteran tree topology. *Syst Entomol* 33:429-450.
5. Lyons-Weiler J, Hoelzer GA & Tausch RJ (1998). Optimal outgroup analysis. *Biol J Linn Soc* 64: 493-511.
6. Maddison WP, Donoghue MJ & Maddison DR (1984). Outgroup analysis and parsimony. *Syst Zool* 33:83-103.
7. Wheeler WC (1990). Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6:363-367.
8. Sanderson MJ & Shaffer HB (2002). Troubleshooting molecular phylogenetic analysis. *Annu Rev Ecol Syst* 33:49-72.
9. Watrous LE & Wheeler QD (1981). The outgroup comparison method of character analysis. *Syst Zool* 30:1-11.
10. Tarrío R, Rodríguez-Trelles F & Ayala F (2000). Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup. The *Grosophila saltans* and *willistoni* groups, a case study. *Mol Phylo Evol* 16: 344-349.
11. Graham SW, Olmstead RG & Barrett SCH (2002). Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Mol Biol Evol* 19: 1769- 1781.
12. Bergsten, J. 2005. A review of long-branch attraction. *Cladistics*, 21: 163-193.
13. Dececchi TA, Narbonne, GM, Greentree and Laflamme M. (2018). Phylogenetic relationships among the Rangeomorpha: the importance of outgroup selection and implications for their diversification. *Can J earth Sci.* 55: 1223-1239.
14. Nasir J (1985). Solanaceae In: Ali SI and Nasir E (eds). Flora of Pakistan, Fascicle 168. Pak. Agric Research council, Islamabad, 61.
15. Jafri SM & El-Gadi (1978). Flora of Libya. *Solanaceae*. 62: 1-38.
16. Migahid AM (1988). Flora of Saudi Arabia. (3<sup>rd</sup> Edition). Cambridge University Press. pp 145-156.
17. Doyle JJ & Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11-15.
18. Rozen S & Skaletsky HJ (2000). Primer 3 on the WWW for general users and for biologist programmers. In: S.Krawetz, S Misener (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Human Press, NJ, USA, pp 365-386.
19. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* 215: 403-410.
20. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA & Wagner L (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31: 28 -33.
21. Corpet F (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16: 10881-10890.
22. Gasteiger EA, Gattiker C, Hoogland I, Ivanyi I, Appel RD & Bairoch A (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31: 3784-3788.
23. Posada D (2008). J Model Test: phylogenetic model averaging. *Mol Biol Evol* 25: 1253-1256.
24. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29: 1969-1973.
25. Sun SM, Yang SH, Golokhvast KS, Le B & Chung G (2016). Reconstructing the Phylogeny of *Capsosiphon fulvescens* (Ulotrichales, Chlorophyta) from Korea

- Based on *rbcL* and 18S rDNA Sequences. *Bio Med Res Inter* 2016: 1-6.
26. Murali S, Rashmi TR & Francis MS (2016). Phylogenetic relationship of selected *Aristolochia* spp. with different generic segregates inferred using *rbcL* and *matK* genes. *J Med Plants* 4: 99-103.
  27. Sarhan S, Hamed F & Al-Youssef W (2016). The *rbcL* gene sequence variations among and within *Prunus* species. *J Agr Sci Tech* 18: 1105-1115.
  28. Chetia P, Phukan BC & Tamang D (2016). *rbcL* gene based molecular phylogenetic studies of certain Pteridophytes of Assam. *Research & Reviews: J Bioinformatics* 3: 18-21.
  29. Florence R, Gey D, Kurihara A, Maggs CA, Martin-Lescanne J, Payri C, Reviers B, Sherwood AR and Le Gall L (2017). Molecular phylogenies support taxonomic revision of three species of *Laurencia* (Rhodomelaceae, Rhodophyta), with the description of a new genus. *Eur J Taxon* 269: 1-19.
  30. Wilberg EW (2015). What's in an outgroup? The impact of outgroup choice on the phylogenetic position of *Thalattosuchia* (Crocodylomorpha) and the origin of *Crocodyliformes*. *Syst Biol* 64: 621-637.
  31. De La Torre-bárcena, JE, Kolokotronis SO, Lee EK, Stevenson DW, Brenner ED, Katari MS, Coruzzi GM & De Salle R (2009). The impact of outgroup choice and missing data on major seed plant phylogenetics using genome-wide EST data. *PLoS One* 4: e5764.
  32. A-Rong L, YanZhou Z, HuiJie Q, WeiFeng S, Murphy RW & ChaoDong Z (2010). Outgroup selection in tree reconstruction: a case study of the family Halictidae (Hymenoptera: Apoidea). *Acta Entomologica Sinica* 53: 192-201.
  33. Kirchberger PC, Sefc KM, Sturmbauer C & Koblmüller S (2014). Outgroup effects on root position and tree topology in the AFLP phylogeny of a rapidly radiating lineage of cichlid fish. *Mol Phylogene Evol* 70: 57-62.
  34. Smarda P, Bureš P, Horová L, Leitch IJ, Mucina L, Pacini E, Lubomír T, Vit G & Rotreklová O (2014). Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc Natl Acad Sci* 111: 4096-4102.
  35. Singh R, Ming R & Yu Q (2016). Comparative analysis of GC content variations in plant genomes. *Trop Plant Biol* 9: 136-149.
  36. Zerui Y, Huang Y, Wenli A, Zheng X, Huang S & Liang L (2019). Sequencing and Structural Analysis of the Complete Chloroplast Genome of the Medicinal Plant *Lycium chinense* Mill. *Plant* 8:87.
  37. Olmstead RG, Sweere JA, Spangler RE, Bohs L & Palmer JD (1999). Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. *Solanaceae IV* 1: 1-137.
  38. Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF & Collier SM (2008). A molecular phylogeny of the Solanaceae. *Taxon* 57:1159-1181.
  39. Martins TR & Barkman TJ (2005). Reconstruction of Solanaceae phylogeny using the nuclear gene SAMT. *Syst Bot* 30: 435-447.
  40. Huelsenbeck JP & Rannala B (2004). Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol* 53: 904-913.
  41. Kim SC, Kim JS, Chase MW, Fay MF & Kim JH (2016). Molecular phylogenetic relationships of Melanthiaceae (Liliales) based on plastid DNA sequences. *Bot J Linn Soc* 181: 567-584.
  42. Simmons M., Pickett KM & Miya M (2004). How meaningful are Bayesian support values? *Mol Biol Evol* 21: 188-19.